ECE 6882 Osama Yousuf

Homework 5 Date: ____________

Problem 1:

$\gamma = 0.9$

$s = \{-1, 1, 2\}$

$a = \{-1, 0, 1\}$

$D = \{(s_0 = 1, a_0 = 1, r_1 = 1, s_1 = 2),$
$(s_1 = 2, a_1 = 0, r_2 = -1, s_2 = 1),$
$(s_2 = 1, a_2 = -1, r_3 = 0, s_3 = -1)\}$

$\phi(s,a) = [a^2 s + as + a]$

$w^0 = 1$

Compute w', w^2, and π^2.

Iteration 1:

Evaluation: $w^- = w^0 = 1$

$Q(s,a) = \phi^T(s,a) w^-$ for $\forall s, a$

$$\Rightarrow Q^{\pi}(s,a) = \begin{array}{c|ccc} & a_0 & a_1 & a_2 \\ s_0 & \phi(-1,-1) & \phi(-1,0) & \phi(-1,1) \\ s_1 & \phi(1,-1) & \phi(1,0) & \phi(1,1) \\ s_2 & \phi(2,-1) & \phi(2,0) & \phi(2,1) \end{array}$$

$$\Rightarrow Q^{\pi}(s,a) = \begin{bmatrix} -1 & 0 & -1 \\ -1 & 0 & 3 \\ -1 & 0 & 5 \end{bmatrix}$$

$$\pi^0 = \begin{array}{c} s=-1 \\ s=1 \\ s=2 \end{array} \begin{bmatrix} \arg\max_a \{-1, \underline{0}, -1\} \\ '' \ \{-1, 0, \underline{3}\} \\ '' \ \{-1, 0, \underline{5}\} \end{bmatrix}$$

$$\Rightarrow \pi^0 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

Date:

Improvement:

$$A = \frac{1}{L} \sum_{i=1}^{n} \phi_t(s_i, a_i)\left[\phi(s_i, a_i) - \gamma \phi(s_{i+1}, \pi(s_{i+1}))\right]^T$$

where, $L = len(D) = 3$

$i=1$: $A = 0$, $b = 0$

$D[1] = (s_0 = 1, a_0 = 1, r_1 = 1, s_1 = 2)$

$\Rightarrow A \mathrel{+}= \phi(1,1) \times [\phi(1,1) - 0.9 \times \phi(2, \pi^0(2))]^T$

$\Rightarrow A \mathrel{+}= \phi(1,1) \times [\phi(1,1) - 0.9 \times \phi(2,1)]^T$

$\Rightarrow A \mathrel{+}= 3(3 - 0.9 \times 5)$

$\Rightarrow \boxed{A = -4.5}$

$b \mathrel{+}= \phi(1,1) \times 1$

$\boxed{b = 3}$

$i=2$:

$D[2] = (s_1 = 2, a_1 = 0, r_2 = -1, s_2 = 1)$

$A \mathrel{+}= \phi(2,0) \times [\phi(2,0) - 0.9 \times \phi(1, \pi^0(1) = 1)]$

$A \mathrel{+}= 0$

$\boxed{A = -4.5}$

$b \mathrel{+}= \phi(2,0) \times -1$

$b \mathrel{+}= 0$

$\boxed{b = 3}$

$i=3$:

$D[3] = (s_2 = 1, a_2 = -1, r_3 = 0, s_3 = -1)$

$A \mathrel{+}= \phi(1,-1) \times [\phi(1,-1) - 0.9 \times \phi(-1, \pi^0(-1) = 0)]$

$A \mathrel{+}= -1 \times [-1 - 0.9 \times 0]$

Date: ____________

$A \mathrel{+}= 1$

$\Rightarrow \boxed{A = -3.5}$

$b \mathrel{+}= \phi(1,-1) \times 0$

$b \mathrel{+}= 0$

$\Rightarrow \boxed{b = 3}$

Normalizing,

$A = -3.5 / 3 = -1.1666$

$b = 3 / 3 = +1$

$w^{+} = A^{-1} b = \frac{-1}{1.1666} \times 1$

$$\boxed{w^{+} = -0.85714}$$

Iteration 2:

Evaluation:

$w^{-} = w^{+} = -0.85714$

$$Q^{\pi}(s,a) = \begin{bmatrix} 0.85714 & 0 & 0.85714 \\ 0.85714 & 0 & -2.57142 \\ 0.85714 & 0 & -4.2857 \end{bmatrix}$$

$$\pi' = \begin{bmatrix} \arg\max_a \; \underline{0.85}, 0, 0.85 \\ \text{"} \quad \underline{0.85}, 0, -2.57 \\ \text{"} \quad \underline{0.85}, 0, -4.28 \end{bmatrix}$$ tie. Prefer -1.

$$\pi' = \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}$$

Date: ____________

Improvement:

$$A = \frac{1}{L}\left(A_1 + A_2 + A_3\right)$$

$$b = \frac{1}{L}\left(b_1 + b_2 + b_3\right)$$

i = 1:

$$A_1 = \overset{3}{\phi(1,1)} \times \left[\phi(1,1) - 0.9 \times \overset{-1}{\phi(2,-1)}\right]$$

$$\boxed{A_1 = 11.7}$$

$$b_1 = \phi(1,1) \times 1 = 3 \times 1$$

$$\boxed{b_1 = 3}$$

i = 2:

$$A_2 = \overset{0}{\phi(2,0)} \times \left[\phi(2,0) - 0.9 \times \phi(1,-1)\right]$$

$$\boxed{A_2 = 0}$$

$$b_2 = \phi(2,0) \times -1$$

$$\boxed{b_2 = 0}$$

i = 3:

$$A_3 = \overset{-1}{\phi(1,-1)} \times \left[\phi(1,-1) - 0.9 \times \overset{-1}{\phi(-1,-1)}\right]$$

$$\boxed{A_3 = 0.1}$$

$$b_3 = \phi(1,-1) \times 0$$

$$\boxed{b_3 = 0}$$

Normalizing,

$$A = (11.7 + 0 + 0.1)/3 = 3.9333$$

$$b = 3/3 = 1$$

$$w^{+} = A^{-1} b = \frac{1}{3.9333} \times 1$$

Date:

$\boxed{w^+ = 0.2542} = w^2$

Finally,

Iteration 3:

Evaluation,

$$w^- = w^+ = w^2 = 0.2542$$

$$Q^\pi(s,a) = \begin{bmatrix} -0.2542 & 0 & -0.2542 \\ -0.2542 & 0 & 0.76 \\ -0.2542 & 0 & 1.271 \end{bmatrix}$$

$$\pi^2 = \begin{bmatrix} \arg\max_a & -0.25, \underline{0}, -0.25 \\ '' & -0.25, 0, \overline{0.76} \\ '' & -0.25, 0, \underline{1.27} \end{bmatrix}$$

$$\pi^2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

Thus,

$$w^1 = -0.8574$$

$$w^2 = 0.2542$$

$$\pi^2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

Date:

Problem 2.

$$\phi(s,a) = \begin{bmatrix} as + a \\ a^2 s \end{bmatrix}_{(k \times 1)} \quad k = 2 \text{ (k basis functions)}$$

$$w^0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}_{(k \times 1)}$$

~~Iteration~~

Evaluation:

$Q^\pi(s,a) = \phi^T(s,a) \times w^-$ for $\forall$ s,a

Recall,

$S = \{-1, 1, 2\}$

$A = \{-1, 0, 1\}$

$Q^\pi(-1,-1) = as + a + a^2 s = -1$

Since $\phi^T(s,a) \times w^-$ is unchanged from Problem 1, the Q-values will be the same. This is because the two basis functions under w^0 give the same basis function as Problem 1.

Thus,

$$Q^\pi(s,a) = \begin{bmatrix} -1 & 0 & -1 \\ -1 & 0 & 3 \\ -1 & 0 & 5 \end{bmatrix}$$

$$\pi^0 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \begin{matrix} s = -1 \\ s = 1 \\ s = 2 \end{matrix}$$

Iteration 1:

Improvement:

$$A = \frac{1}{L}(A_1 + A_2 + A_3)$$

$$b = \frac{1}{L}(b_1 + b_2 + b_3)$$

i = 1:

$$A_1 = \phi(1,1) \times [\phi(1,1) - 0.9\,\phi(2,1)]^T$$

$$\phi(1,1) = \begin{bmatrix} 1+1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$\phi(2,1) = \begin{bmatrix} 2+1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

$$\cancel{A_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}_{(k\times 1)} \times \begin{bmatrix} 3 & 2 \end{bmatrix}_{(1\times k)}}$$

$$\cancel{\boxed{A_1 = \begin{bmatrix} 6 & 4 \\ 3 & 2 \end{bmatrix}}}$$

$$\Rightarrow A_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \times \left(\begin{bmatrix} 2 \\ 1 \end{bmatrix} - 0.9 \times \begin{bmatrix} 3 \\ 2 \end{bmatrix} \right)^T$$

$$A_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \times \begin{bmatrix} -0.7 & -0.8 \end{bmatrix}$$

$$\boxed{A_1 = \begin{bmatrix} -1.4 & -1.6 \\ -0.7 & -0.8 \end{bmatrix}}$$

$$b_1 = \phi(1,1) \times 1$$

$$\boxed{b_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}}$$

Date:

$i = 2:$

$$A_2 = \phi(2,0) \times [\phi(2,0) - 0.9\phi(1,1)]^T$$

where,

$$\phi(2,0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \phi(1,1) = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$A_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \times \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix} - 0.9 \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right)^T$$

$$\boxed{A_2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}}$$

$$b_2 = \phi(2,0) \times -1$$

$$\boxed{b_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}}$$

$i = 3:$

$$A_3 = \phi(1,-1) \times [\phi(1,-1) - 0.9\phi(-1,0)]^T$$

$$A_3 = \begin{bmatrix} -1-1 \\ 1 \end{bmatrix} \times \left(\begin{bmatrix} -2 \\ 1 \end{bmatrix} - 0.9 \times \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right)^T$$

$$A_3 = \begin{bmatrix} -2 \\ 1 \end{bmatrix} \times \begin{bmatrix} -2 & 1 \end{bmatrix}$$

$$\boxed{A_3 = \begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix}}$$

$$b_3 = \phi(1,-1) \times 0$$

$$\boxed{b_3 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}}$$

Finally,

$$A = \frac{1}{3}\left(\begin{bmatrix} -1.4 & -1.6 \\ -0.7 & -0.8 \end{bmatrix} + 0 + \begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix}\right)$$

$$A = \begin{bmatrix} 0.866 & -1.2 \\ -0.9 & 0.066 \end{bmatrix}$$

and, $$b = \frac{1}{3}\left(\begin{bmatrix} 2 \\ 1 \end{bmatrix}\right)$$

$$b = \begin{bmatrix} 0.666 \\ 0.333 \end{bmatrix}$$

$$\Rightarrow \quad w' = A^{-1} b$$

$$A^{-1} = \frac{1}{(0.066 \times 0.866) - (1.2 \times 0.9)} \times \begin{bmatrix} 0.066 & 1.2 \\ 0.9 & 0.866 \end{bmatrix}$$

$$A^{-1} = -\frac{1}{1.022} \times \begin{bmatrix} 0.066 & 1.2 \\ 0.9 & 0.866 \end{bmatrix}$$

$$\Rightarrow \quad w' = \frac{-1}{1.022} \times \begin{bmatrix} 0.066 \times 0.666 + 1.2 \times 0.333 \\ 0.9 \times 0.666 + 0.866 \times 0.333 \end{bmatrix}$$

$$\Rightarrow \quad \boxed{w' = \begin{bmatrix} -0.434 \\ -0.866 \end{bmatrix}}$$

<u>Evaluation for π'</u>:

$$Q^{\pi}(s,a) = \phi^{\pi}(s,a) \times w'$$

$$\Rightarrow \quad Q^{\pi}(-1,-1) = \begin{bmatrix} as+a & a^2 s \end{bmatrix}_{1 \times k} \times \begin{bmatrix} -0.434 \\ -0.866 \end{bmatrix}_{k \times 1}$$

$$= -0.434(-1 \cdot -1 + -1) - 0.866 \times -1$$

$$\boxed{Q^{\pi}(-1,-1) = +0.866}$$

Date:

$$\boxed{Q^{\pi}(-1,0) = 0}$$

$$Q^{\pi}(-1,1) = \begin{bmatrix} as+a \\ a^2 s \end{bmatrix}^T \begin{bmatrix} -0.434 \\ -0.866 \end{bmatrix}$$

$$= \begin{bmatrix} -1+1 & -1 \end{bmatrix} \begin{bmatrix} -0.434 \\ -0.866 \end{bmatrix}$$

$$\boxed{Q^{\pi}(-1,1) = +0.866}$$

Repeat for other states.
Calculations done in code.

$$Q^{\pi}(s,a) = \begin{array}{c} s=-1 \\ s=1 \\ s=2 \\ \\ \end{array} \begin{array}{ccc} \begin{bmatrix} \underline{0.866} & 0 & 0.866 \\ 0.002 & \underline{0} & -1.734 \\ -0.434 & \underline{0} & -3.034 \end{bmatrix} \\ \;\; a=-1 \quad\quad a=0 \quad\quad a=1 \end{array}$$

* underlined q-values indicate optimal action
* for $s=-1$, tie is broken by prefering $a=-1$
* for $s=1$, $a=-1$ & $a=0 \approx 0$, tie is broken by preffering $a=0$.

$$\Rightarrow \quad \pi' = \begin{array}{c} \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} \end{array} \begin{array}{l} s=-1 \\ s=1 \\ s=2 \end{array}$$

Iteration 2:

Improvement:

i = 1:

$$A_1 = \phi(1,1) \times [\phi(1,1) - 0.9 \times \phi(2,0)]^T$$

$$= \begin{bmatrix} 2 \\ 1 \end{bmatrix} \times \left(\begin{bmatrix} 2 \\ 1 \end{bmatrix} - 0.9 \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right)^T$$

$$A_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \end{bmatrix}$$

$$\boxed{A_1 = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}}$$

$$b_1 = \phi(1,1) \times 1 = \boxed{\begin{bmatrix} 2 \\ 1 \end{bmatrix}}$$

i = 2:

$$A_2 = \phi(2,0) \times [\phi(2,0) - 0.9 \phi(1,0)]^T$$

$$\boxed{A_2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}}$$

$$b_2 = \phi(2,0) \times -1$$

$$\boxed{b_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}}$$

i = 3:

$$A_3 = \phi(1,-1) \times [\phi(1,-1) - 0.9 \phi(-1,-1)]^T$$

$$= \begin{bmatrix} -2 \\ 1 \end{bmatrix} \times \left(\begin{bmatrix} -2 \\ 1 \end{bmatrix} - 0.9 \times \begin{bmatrix} -1 \\ -1 \end{bmatrix} \right)^T$$

$$= \begin{bmatrix} -2 \\ 1 \end{bmatrix} \times \begin{bmatrix} -2 \\ 1.9 \end{bmatrix}^T = \begin{bmatrix} -2 \\ 1 \end{bmatrix} \begin{bmatrix} -2 & 1.9 \end{bmatrix}$$

$$\boxed{A_3 = \begin{bmatrix} 4 & -3.8 \\ -2 & 1.9 \end{bmatrix}}$$

Date:

$b_3 = \phi(1,-1) \times 0$

$$b_3 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Finally,

$$A = \frac{1}{3}(A_1+A_2+A_3)$$

$$A = \begin{bmatrix} 2.666 & -0.6 \\ 0 & 0.966 \end{bmatrix}$$

$$b = \begin{bmatrix} 0.666 \\ 0.333 \end{bmatrix}$$

$$A^{-1} = \begin{bmatrix} 0.375 & 0.232 \\ 0 & 1.034 \end{bmatrix}$$

$$\Rightarrow w^2 = A^{-1}b = \begin{bmatrix} 0.375 \times 0.666 + 0.232 \times 0.333 \\ 0 + 1.034 \times 0.333 \end{bmatrix}$$

$$\Rightarrow \boxed{w^2 = \begin{bmatrix} 0.327 \\ 0.344 \end{bmatrix}}$$

Evaluation for π^2:

$$Q^{\pi}(s,a) = \phi^T(s,a) \times \begin{bmatrix} 0.327 \\ 0.344 \end{bmatrix}$$

where,

$$\phi^T(s,a) = \begin{bmatrix} as+a & a^2 s \end{bmatrix}$$

$$\Rightarrow Q^{\pi}(s,a) = \begin{array}{c} s=-1 \\ s=1 \\ s=2 \\ \\ \end{array} \begin{array}{ccc} \left[\begin{array}{ccc} -0.344 & 0 & -0.344 \\ -0.31 & 0 & 1 \\ -0.29 & 0 & 1.6 \end{array}\right] \\ a=-1 \quad a=0 \quad a=1 \end{array}$$

Date:

* undelined values indicate argmax.

$$\Rightarrow \boxed{\pi^2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}}$$

Thus,

$$w^1 = \begin{bmatrix} -0.434 \\ -0.866 \end{bmatrix}$$

$$w^2 = \begin{bmatrix} 0.327 \\ 0.344 \end{bmatrix}$$

$$\pi^2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

→ The final policy in Problem 1 & 2 is the same, but this can not be true in the general case.

→ For example, having $\phi(s,a) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ in this problem would lead to different final policies.

× ——— × ——— ×